Reading Assignment-1
Definitions and Concepts

1. **Machine Learning:** ML is a set of methods that computers use to make and improve predictions or behaviors based on data. The machine learning algorithm learns a model by estimating parameters (like weights) or learning structures (like trees). The algorithm is guided by a score or loss function that is minimized.
   The predictions will be made based on supervised, unsupervised learning and reinforcement learning.
   Supervised learning is something that has a reference of past data.
   For Example: Predicting the value of a house basing on the past housing prices. The goal of supervised learning is to learn a predictive model that maps features of the data (e.g. house size, location, floor type, ...) to an output (e.g. house price). If the output is categorical, the task is called classification, and if it is numerical, it is called regression.
   While Unsupervised learning is where we do not have a specific outcome of interest, but want to find the clusters of data points.
   Reinforcement learning is where an agent learns to optimize a certain reward by acting in an environment(for example computer playing Tetris).
2. **Clustering**: Clustering is an unsupervised model of machine learning where we learn from data properties and labelling of group of points. It is the task of dividing the data points or population into several number of groups such that data points in the same groups are more similar to other data points in the same group and different to the data points in other groups.
   Different types of clustering is as follows :
   a. Density based clustering method
   b. Hierarchical based clustering method
   c. Partitioning method
   d. Grid based method
   Example of Clustering is Kmeans Clustering, Gaussian mixtures, Agglomorative clustering, Spectral clustering etc..
3. **K-means Clustering:** This algorithm finds a pre-determined no. of clusters within an unlabeled multidimensional dataset. It partition the observations into 'k' clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster and where k is the number of clusters. The number of clusters can be determined using elbow method. It determines points in such a way that they satisfy the below assumptions:
   a. The center of cluster is the average of all points belonging to that cluster
   Each point present in a cluster is closer to its own cluster center than to any other cluster centers. The k-means algorithm divides a set of N samples X into K disjoint clusters C, each described by the mean μj of the samples in the cluster. The means are commonly called the cluster "centroids"; note that they are not, in general, points from X, although they live in the same space.

The K-means algorithm aims to choose centroids that minimise the **inertia**, or **within-cluster sum-of-squares criterion**:

$$\sum_{i=0}^{n} \min_{\mu_j \epsilon C} \left(||x_i - \mu_j||^2 \right)$$

Example of Kmeans Clustering is Expectation-Maximization.

4. **Expectation-Maximization** is a powerful algorithm that is based on kmeans clustering. The approach of E-M algorithm consist of a procedure that guesses some cluster centers and it repeats until they converge.
   E-step is the Expectation step that assigns a point to the nearest cluster center. This step involves updating our expectation to which the cluster belongs to.
   M-step is the Maximization step where it sets the cluster centers to the mean. This step maximizes some fitness function that defines the cluster center location.
   Example:

5. **Interpretability:** There are different definitions for interpretability. It can be defined as the degree to which a human can understand the cause of a decision. or can consistently predict the model's result. When the interpretability of a machine learning model is high, then it makes it easy for human to comprehend the reason why certain decisions or predictions have been made. An algorithm that trains a model that produces the predictions. Each step can be evaluated in terms of transparency or interpretability. Doshi-Velez and Kim proposed 3 main levels of evaluation interpretability
   a. Application level evaluation(real task)
   b. Human level evaluation(simple task)
   c. Function level evaluation(proxy task)
   For example, the probability that a customer will churn or how effective some drug will be for a patient. The need for interpretability arises from an incompleteness in problem formalization), which means that for certain problems or tasks it is not enough to get the prediction. The model must also explain how it came to the prediction (the **why**), because a correct prediction only partially solves your original problem

6. **Feature Selection:** It is the process of reducing the input variable(features) by removing less important input variables before we feed into the model.
   There are some things that are to be performed in feature selection such as Keeping relevant features, removing irrelevant and redundant features.
   Some benefits of this process is as follows: 1. Reduces overfitting which leads to less redundant data, Improves accuracy by removing useless features and less features makes algorithm complexity easier and faster and thus reducing Training time and complexity.

7. **Techniques of Feature selection**

There are different Techniques to perform feature selection. They are

A. **Filter method**: Here the model starts with all the features but selects best feature subset based on measures of statistics such as Pearson's correlation, Linear Discriminant Analysis (LDA), ANOVA, Chi-square, etc. Types of Filter methods is Univariate and multivariate which is based on the number of features used to filter them. These are classified further as a) Basic Filter method b) Correlation filter method c) Statistical and Ranking method

B. **Wrapper method** : This method starts by creating a number of machine learning models with different subsets of input features and select these combinations of features that result in the best performing model according to a performance metric

C. **Embedded method** : This method combines the qualities of filter and wrapper methods. It performs feature selection during the model training, which is why we call them embedded methods. These are implemented by algorithms that have their own built-in feature selection methods.

D. **Hybrid method** This method combines various ways to get best possible feature subset. It combines pieces of wrapper methods and embedded methods.

8. **Terminologies in ML:**
   a. **Algorithm** is a set of rules that a machine follows to achieve a particular goal[2]. An algorithm can be considered as a recipe that defines the inputs, the output and all the steps needed to get from the inputs to the output.
   b. **Learner or ML Algorithm** is the program used to learn a machine learning model from data. Another name is "inducer" (e.g. "tree inducer").
   c. **Dataset** is a table with data from which the machine learns. It contains the features and the target to predict. When used to induce a model, the dataset is called training data.
   d. **Instance** is a row in the dataset. It is also called as (data) point, example, observation. It consists of the feature values $x^{(i)}$ and, if known, the target outcome y_i.
   e. **Features are used as inputs for ML prediction or classification. These are columns in the input dataset** The matrix with all features is called X and $x^{(i)}$ for a single instance. The vector of a single feature for all instances is x_j and the value for the feature j and instance i is $x_j^{(i)}$.
   f. **Target** is the information the machine learns to predict. In mathematical formulas, the target is usually called y or y_i for a single instance.
   g. **ML Task** is the combination of a dataset with features and a target. Depending on the type of the target, the task can be for example classification, regression, survival analysis, clustering, or outlier detection.
   h. **Prediction** is something that the machine learning model "guesses" and tells what the target value should be based on the given features. The model prediction can be denoted by $f(x^{(i)})$ or y.